HOST ORDER EFFECT ON VIRUS SEGMENTATION

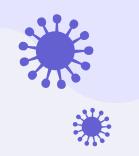EEB313: Juwon (Lucia) Park, Yuehan (Anna) Li , Cindy Yu

WHY STUDY THIS?

- Segmented viruses pose a significant risk
- Understanding the role of host order helps us identify patterns that influence susceptibility
- This research can help with future viral outbreaks

HYPOTHESIS AND PREDICTION

BACKGROUND

VIRUS SEGMENTATION

- A viral genome is sometimes split into parts (segmentation)

- Enables virus to respond dynamically to host environments

WHY HOST ORDER?

- Host order with complex genomes may drive evolution of segmented viral genomes

- E.g. Sialic acid receptors

MAIN HYPOTHESIS

- Host order is a significant driver of whether mammal species are more likely to be affected by segmented viruses

- Segmented viruses are uniquely equipped to adapt dynamically to the host

Key question: Does host order act as a **primary** determinant of susceptibility?

PREDICTION

1

MODEL 1

Other predictors

2

MODEL 2

Other predictors
AND
Host order

Model 2 (with host order) will have a **lower AIC**!
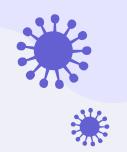
DATA

DATA DESCRIPTION

- CLOVER database
    - Mammal_viruses_Associations
    - HP3_Virus dataset
- Mammal_viruses_Associations
    - Information on mammalian host order and associated viruses
- HP3_Virus
    - Predictors

Data Cleaning and Integration

- Virus name standardized
  - Lowercase
  - Underscore
- DNA virus excluded
  - No segmented DNA virus
- Host order
  - Fewer than 40 rows removed
  - Remove NAs

OTHER PREDICTORS

1 GENOME LENGTH

2 CYTOPLASMIC REPLICATION

3 ENVELOPE

4 SINGLE/DOUBLE STRAND

5 ZOONOTIC

WHY LOGISTIC REGRESSION?

Generalized linear model with **BINARY** (Bernoulli, binomial) data:

PREDICTORS $\updownarrow$ = MEAN OF RESPONSE $\updownarrow$

| vCytoReplicTF |
| vGenomeAveLength |
| vSSoDS |
| IsZoonotic |
| vEnvelope |

$\updownarrow$ =

vSegmentedTF
FALSE
FALSE
TRUE
TRUE
TRUE
TRUE

Eg. a one-unit change in X_i changes the log-odds of Y (being true) by _____

LOGISTIC REGRESSION

```r
```{r}
model = glm(vSegmentedTF ~ vGenomeAveLength*vEnvelope  + vSSoDS*IsZoonotic + vCytoReplicTF + vEnvelope,
family = 'binomial', data = df_hostremoved)

summary(model)

model2 = glm(vSegmentedTF ~ HostOrder + vGenomeAveLength*vEnvelope  + vSSoDS*IsZoonotic + vCytoReplicTF +
vEnvelope, family = 'binomial', data = df_hostremoved)

summary(model2)
```
```

Traits **interact**
- Mechanisms not entirely clear:
  - Genome size constraints, coding for envelopes, rapid mutation rates
Must avoid collinearity

Coefficients: (3 not defined because of singularities)

LOGISTIC REGRESSION RESULTS

```
Call:
glm(formula = vSegmentedTF ~ vGenomeAveLength * vEnvelope + vSSoDS *
    IsZoonotic + vCytoReplicTF + vEnvelope, family = "binomial",
    data = df_hostremoved)

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                -2.287e+01  1.991e+02  -0.115    0.909
vGenomeAveLength            5.069e-04  5.337e-04   0.950    0.342
vEnvelope                   2.220e+01  3.511e+02   0.063    0.950
vSSoDS                     -2.170e+01  3.511e+02  -0.062    0.951
IsZoonotic                  1.976e+01  3.152e+02   0.063    0.950
vCytoReplicTFTRUE           1.819e+01  1.988e+02   0.092    0.927
vGenomeAveLength:vEnvelope -4.702e-04  5.338e-04  -0.881    0.378
vSSoDS:IsZoonotic          -1.734e+01  3.152e+02  -0.055    0.956

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 21234.0  on 19600  degrees of freedom
Residual deviance:  7027.1  on 19593  degrees of freedom
AIC: 7043.1

Number of Fisher Scoring iterations: 19
```

```
Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                -2.390e+01  3.147e+02  -0.076   0.9395
HostOrdercarnivora          1.789e+00  1.194e-01  14.977  < 2e-16 ***
HostOrderchiroptera        -2.242e+00  2.211e-01 -10.139  < 2e-16 ***
HostOrderdidelphimorphia   -2.064e+01  4.575e+03  -0.005   0.9964
HostOrderdiprotodontia     -1.123e+00  4.766e-01  -2.356   0.0184 *
HostOrderlagomorpha        -1.166e+00  7.568e-01  -1.541   0.1233
HostOrderperissodactyla    -8.081e-01  1.207e-01  -6.697 2.13e-11 ***
HostOrderprimates          -8.287e-01  8.274e-02 -10.016  < 2e-16 ***
HostOrderrodentia          -3.059e+00  2.746e-01 -11.137  < 2e-16 ***
vGenomeAveLength            5.211e-04  3.241e-04   1.608   0.1079
vEnvelope                   2.373e+01  5.625e+02   0.042   0.9664
vSSoDS                     -2.301e+01  5.626e+02  -0.041   0.9674
IsZoonotic                  2.154e+01  4.885e+02   0.044   0.9648
vCytoReplicTFTRUE           1.933e+01  3.146e+02   0.061   0.9510
vGenomeAveLength:vEnvelope -4.795e-04  3.242e-04  -1.479   0.1391
vSSoDS:IsZoonotic          -1.897e+01  4.885e+02  -0.039   0.9690
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 21234  on 19600  degrees of freedom
Residual deviance:  5915  on 19585  degrees of freedom
AIC: 5947

Number of Fisher Scoring iterations: 20
```

Host orders are sometimes significant!

CONCLUSION

"Does host order act as a **primary** determinant of susceptibility to segmented viruses?"

"Sometimes!"

Chiroptera (bats), carnivora, primates, rodents, perissodactyla (odd-toed hoofed)

```
Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -2.390e+01  3.147e+02  -0.076   0.9395
HostOrdercarnivora        1.789e+00  1.194e-01  14.977  < 2e-16 ***
HostOrderchiroptera      -2.242e+00  2.211e-01 -10.139  < 2e-16 ***
HostOrderdidelphimorphia -2.064e+01  4.575e+03  -0.005   0.9964
HostOrderdiprotodontia   -1.123e+00  4.766e-01  -2.356   0.0184 *
HostOrderlagomorpha      -1.166e+00  7.568e-01  -1.541   0.1233
HostOrderperissodactyla  -8.081e-01  1.207e-01  -6.697 2.13e-11 ***
HostOrderprimates        -8.287e-01  8.274e-02 -10.016  < 2e-16 ***
HostOrderrodentia        -3.059e+00  2.746e-01 -11.137  < 2e-16 ***
vGenomeAveLength          5.211e-04  3.241e-04   1.608   0.1079
vEnvelope                 2.373e+01  5.625e+02   0.042   0.9664
vSSoDS                   -2.301e+01  5.626e+02  -0.041   0.9674
IsZoonotic                2.154e+01  4.885e+02   0.044   0.9648
vCytoReplicTFTRUE         1.933e+01  3.146e+02   0.061   0.9510
vGenomeAveLength:vEnvelope -4.795e-04 3.242e-04 -1.479   0.1391
vSSoDS:IsZoonotic        -1.897e+01  4.885e+02  -0.039   0.9690
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 21234  on 19600  degrees of freedom
Residual deviance:  5915  on 19585  degrees of freedom
AIC: 5947

Number of Fisher Scoring iterations: 20
```

Thank you!

Questions?